

## The NASA Exoplanet Science Institute Archives: KOA and NStED

G. B. Berriman, D. Ciardi, M. Abajian, T. Barlow, G. Bryden,  
K. von Braun, J. Good, S. Kane, M. Kong, A. Laity, M. Lynn,  
D. Mc Elroy, P. Plavchan, S. Ramirez, M. Schmitz, J. Stauffer,  
P. Wyatt, and A. Zhang

*Infrared Processing and Analysis Center, California Institute of  
Technology, Pasadena, CA, USA*

R. Goodrich, J. Mader, H. Tran, and M. Tsubota

*W. M. Keck Observatory, Waimea, HI, USA*

A. Beekley, S. Berukoff, B. Chan, C. Lau, M. Regelson, M. Saucedo,  
and M. Swain

*Raytheon Corporation, Pasadena, CA, USA*

**Abstract.** The NASA Exoplanet Science Institute (NExSci) maintains a series of archival services in support of NASA's planet finding and characterization goals. Two of the larger archival services at NExSci are the Keck Observatory Archive (KOA) and the NASA Star and Exoplanet Database (NStED). KOA, a collaboration between the W. M. Keck Observatory and NExSci, serves raw data from the High Resolution Echelle Spectrograph (HIRES) and extracted spectral browse products. As of June 2009, KOA hosts over 28 million files ( 4.7 TB) from over 2,000 nights. In Spring 2010, it will begin to serve data from the Near-Infrared Echelle Spectrograph (NIRSPEC). NStED is a general purpose archive with the aim of providing support for NASA's planet finding and characterization goals, and stellar astrophysics. There are two principal components of NStED: a database of (currently) all known exoplanets, and images; and an archive dedicated to high precision photometric surveys for transiting exoplanets. NStED is the US portal to the CNES mission CoRoT, the first space mission dedicated to the discovery and characterization of exoplanets. These archives share a common software and hardware architecture with the NASA/IPAC Infrared Science Archive (IRSA). The software architecture consists of standalone utilities that perform generic query and retrieval functions. They are called through program interfaces and plugged together to form applications through a simple executive library.

## 1 Introduction

The NASA Exoplanet Science Institute (NExSci) Science is the science operations and analysis center for NASA's Exoplanet Exploration Program. It is part of the greater Infrared Processing and Analysis Center (IPAC) at Caltech. This talk will describe two archives with large exoplanet content housed at NExSci: the Keck Observatory Archive (KOA), a collaboration between NExSci and the W. M. Keck Observatory and which opened for business in July 2006 (<https://>

`koa.ipac.caltech.edu`); and the NASA Star and Exoplanet Database (NStED), which opened for business in July 2007 (<http://nsted.ipac.caltech.edu>).

## 2 Building KOA and NStED on the IPAC Archive Heritage

KOA and NStED take full advantage of the archive expertise at IPAC. They share a common hardware and software architecture with the NASA/IPAC Infrared Science Archive (IRSA), which has since 1999 been the archive for NASA's infrared data sets. NStED and KOA inherit from IRSA a science information system for processing user queries; a configuration management system and code repository; testing methodology; user support tools, and a mass-storage and server architecture.

NStED is a close stellar analog of the NASA Extragalactic Database (NED), a knowledge base for extragalactic astronomy. NED's relational database design and name resolver software were adapted in the development of NStED. NED's expertise in cross-identifications was exploited to develop a list of catalog cross-identifications for every star in NStED. Finally, NStED uses the same reference coding for tracking the literature as used by NED and by the SAO/NASA Astrophysics Data System.

## 3 A Common Software Architecture with IRSA

IRSA uses a component based architecture, the Infrared Science Information System (ISIS), designed to enable strong re-use and adaptation. ISIS is optimized for astronomical spatial searches and complex, general data and metadata queries regardless of wavelength and type of mission. To date, it has successfully processed over 25 million queries and more than 100 TB of data have been downloaded through it. In this architecture, each component is a module with a standard interface that communicates with other components and fulfills one general function, such as performing coordinate transformations, generating HTML return pages, or filtering tabular information. The individual modules are stand-alone portable ANSI-C tools. They are plugged together to form user applications, generally CGI-based, with new modules developed as needed, and they are controlled by an executive library. This library starts components as child services and parses return values. Thus KOA and NStED have been able to inherit all the functionality available in ISIS an extended to meet their particular needs with the development of complex proprietary data protection, automated ingestion of data and dynamical generation of interfaces from configuration files (over 150 in the case of NStED).

## 4 Contents of the Keck Observatory Archive

KOA archives all raw data from the High Resolution Echelle Spectrograph (HIRES) from 1994, when the instrument was commissioned, to date. The physical archive is based at NExScI. Raw science and calibration data measured between 1994 and 2004, when the instrument used a single-chip CCD, were recovered from magnetic tape at WKMO and transferred to NEXScI. In

2004, the instrument was upgraded to use a three-chip CCD, and now the raw data are electronically ingested by the afternoon following the observations. PIs have proprietary access to their data, including metadata, for at least 18 months after date of observations. Data are protected per CCD - each CCD recorded in a data file may have a different release date. As of November 2009, data from 678 programs over 1,943 nights are public (representing 200,000+ science CCDs). Altogether, there are 2,218 nights of HIRES data in the archive, containing 87,685 science files and 135,136 calibration files. These data have a volume of 3.1 TB. The archive is in active use by astronomers, and there have been 130,000 queries to it. Bonifacio et al. (2009) have used data from KOA in their study of the chemical history of Galactic metal-poor stars.

#### 4.1 Extracted HIRES Spectra at KOA

One of the most useful products available in KOA are extracted browse spectra. KOA developed an automated pipeline for the extraction of 1-D browse spectra from the raw science and calibration files. It is an upgrade to an existing data reduction package, *makee*, and offers extended wavelength calibration to 9800 Å, and improvements to cosmic-ray masking, trace-finding and slit-finding routines.

The extracted spectra data product contains a 1-D spectrum per order per CCD and an extraction grade that was validated manually, and diagnostic data for each order including signal-to-noise, traces and profiles, preview images and the reduction logs. Altogether, spectra have been extracted for 64,343 observations (31 million files), and their volume is 1.85 TB.

### 5 NASA Star and Exoplanet Database (NStED)

NStED is a database of technical information for nearby stars and exoplanets. There are two components to its scientific content: a database of stellar and exoplanet parameters and associated data, and an archive of light curves from transit survey projects.

#### 5.1 Star and Exoplanet Services

NStED houses a rich database of published and derived parameters and associated data of nearby stars and exoplanets, organized to support complex and detailed searches on potential and known exoplanet stellar hosts. All data are validated and traced to their source by the NStED science staff. Currently, there are data on 145,000 stars, and on all known exoplanets. Published parameters include positions, distances, photometry, colors, multiplicity, orbital parameters etc. Parameters derived from published information use algorithms from the literature, and like the data, are traced to the source papers. Examples are habitable zone sizes, predicted radial velocity wobbles, luminosities, etc. Finally, NStED has ingested associated data such as published light curves, images and radial velocity curves.

#### 5.2 Exoplanet Transit Surveys

NStED hosts an archive of over 250,000 light curves from exoplanet transit surveys, accessible from a dedicated interface. Table 1 summarizes this content.

Table 1. Summary of Exoplanet Survey Content in NStED.

Survey	# Stars	Filter	Time Span(d)	# Epochs
CoRoT Exoplanet	28,078	White 3-color- prism	20–150	3,500–370,000
CoRoT-Seismology	30	White	20–150	50,000–420,000
TrES-Lyr1	25,947	r, R	75	~ 15,000
KELT-Praesepe	66,637	R <sub>K</sub>	73	~ 3,000
NGC 2301	3,961	R	14	~ 150
NGC 3201	58,666	V, I	700	~ 120
M 10	43,930	V, I	500	~ 50
M 12	32,378	V, I	500	~ 50

### 5.3 U. S. CoRoT Data Center

Under an agreement between NASA and CNES, NStED is the U.S. portal to the science data from the CONvection, ROTation and planetary Transits (CoRoT) mission. NStED offers astronomers access to the public CoRoT light curves, and is developing light curve analysis tools to support analysis of these data. The first tool, a periodogram service, will be available in Spring 2010.

## 6 The Near Future for KOA and NStED

In Spring 2010, KOA will begin to serve 1,000 nights of observations (250,000 science and calibration files) measured with the Near Infrared Echelle Spectrograph (NIRSPEC) from 1999 to date. New data will be ingested automatically as they are obtained.

One of NStED’s major goals in the next year is to release light curve analysis tools, beginning with the periodogram tool described above. It will release an exoplanet transit ephemeris service (based on Greg Laughlins Transit Search.org). It will serve new data sets, including CoRoT data sets as released, M2K spectra (spectroscopic survey of late-type stars for planets), and exoplanet spectra measured with the Spitzer Space Telescope. It anticipates augmenting its exoplanet transit survey content by 300,000 light curves from five transit surveys.

**Acknowledgments.** KOA and NStED are funded by the National Aeronautics and Space Administration (NASA). We wish to thank data providers who have generously contributed observations to NStED.

## References

Bonifacio, P. et. al. 2009, A&A, 501, 519.